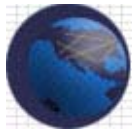## Integrative Cancer Research Special Interest Group Teleconference

Informatics for Proteomics SIG Meeting Minutes

Date, Time & Location:	May 11, 2004 1:00 – 2:00 EDT
Attendees:	Patrick McConnell – Duke (funded developer) Michael Ochs – Fox Chase (funded developer) Tom Moloshok – Fox Chase (funded developer) Jack London – Thomas Jefferson (funded developer) Devjani Chatterjee - Thomas Jefferson (funded developer) Lianhong Tang – Vanderbilt (funded participant) Rakesh Nagarajan- Wash U (funded developer) Steve Enkemann – Moffitt (funded adopter) David Jewell – Dartmouth (funded developer) Subha Madhavan– NCI David Kane – SRA/NCI Claire Zhu – BAH Juli Klemm - BAH
Introduction:	**Roll-call, open meeting, review meeting goals** - Establish goals and priorities for this SIG - For match making purposes, Developers and Adopters will be asked to give a brief statement of their capabilities and interests, respectively - Identify and define additional research

Overview Discussion:	**Review goals and objectives of Data Analysis and Statistical Methods SIG**
	
	Scope: Tools, data and technologies that are relevant to manipulation of proteomics data.
	
	Open Discussion
	
	- The group discussed proteomics technologies in use at the cancer centers
	
	    o Fox Chase: 2D gels are used to separate proteins from fractioned samples on multiple gels run under different conditions to identify proteins involved in various stages of colon cancer, gastro-intestinal cancer and ovarian cancer. MALDI-TOF investigation of serum proteins is being investigated.
	
	    o NCI: Lysate protein arrays are being made for NCI 60 cancer cell lines and are used for antibody screening
	
	    o MALDI-TOF is used at Duke.
	
	    o Other caBIG centers are using SELDI-TOF as well (e.g. MD Anderson)
	
	    o Current proteomics efforts seem to focus on protein identification and quantification, and are mostly applied to the problems of biomarker identification. Post-translational modifications are a lower priority.
	
	    o In general, application of TOF technology is focused on biomarkers while the application of 2D gels is focused on mechanistic questions.
	
	- Data integration
	
	    o Vanderbilt: Linking protein expression data with transcript expression data was cited as a key interest. Accession numbers from databases such as LocusLink, UniGene and Refseq are commonly used to map mRNAs to

proteins.

- o Wash U: Information also needs to be linked back with sample (tissue) data. This is being accomplished at Wash U by linking back to a centralized tissue bank database.

- The group discussed emerging standards for proteomics data

  - o Rakesh (Wash U) has compiled a list of some of the emerging proteomics standards:

    - mzXML format suggested by Institute for System Biology: http://sashimi.sourceforge.net/software.html. Schema, converters, some viewers, some utility programs
    - XML for spectra by gaml.org (Generalized Analytical Markup Language) www.gaml.org/Documentation/XML: Analytical Archive Format.pdf www.gaml.org/Documentation/presentation.pdf: presentation on the use of XML to represent diverse type of spectra. Schema but no software
    - HUPO **Proteomics Standards** Initiative http://psidev.sourceforge.net/
    - PEDRo http://pedro.man.ac.uk/home.shtml

  - o MIAPE (mirroring MIAME) being developed by HUPO may be emerging as the lead in this area. Lianhong (Vanderbilt) is involved in this effort.

    Version available online.
    - http://psidev.sourceforge.net/mi/xml/doc/user/

    Tools used for proteomics standard.
    - http://psidev.sourceforge.net/mi/xml/doc/user/ - tools

    The following database supports proteomics standard
    - http://psidev.sourceforge.net/mi/xml/doc/user/ - submission

    Paper discussing this issue:
    - http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubM ed&list_uids=14755292&dopt=Abstract

High-level review of Center interests in this SIG	**Developers** Fox Chase: Proteomics LIMS - Designed for managing data from a 2D gel + mass spectrometry workflow. Built upon and extended from caLIMS. Added tools for collaborative project management. - Current status is Alpha version 1. Looking for adopters with different data types. Also seeking developers to collaborate. Hoping to build into a multi-sites/multi-developers project. - It is possible to adopt the LIMS with different workflows. This is done by using a multi-tiered architecdture and a modular structure with a presentation layer that can be customized at the individual labs. - Ideal approach to LIMS design is to work closely with individuals in the lab to
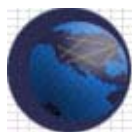
gain a detailed understanding of the workflow and of their processes.

- Data is private until it is "published" by the PI.

<u>Duke</u>: RProteomics and RMetabolomics
http://www.dbsr.duke.edu/research/softwaredev/libraries/r/rproteomics/
- Data analysis tools for MALDI-TOF and SELDI-TOF, e.g. peak picking, clustering. No data management capability. Current status is Alpha release, internal use. Beta release is planned in 6 months.

- Setting up a new 2D gel workflow. Looking to adopt LIMS

<u>WashU:</u> Developed for LC mass spec and regular mass spec workflows.

- Java client-based application

- Would be willing to share expertise they have in the area of proteomics LIMS to support the Fox Chase effort.

<u>Vanderbilt:</u> Has a database that allows linking of mass spec data to tissue and patient data.

<u>NCI</u>: AbMINER tool is a repository of antibody screen data across intramural NCI labs. Possible caBIG data source.

<u>Dartmouth</u>: Q5 tool: http://www.cs.dartmouth.edu/~donaldlab/Software/ Analysis of SELDI-TOF data

Adopters

<u>Thomas Jefferson</u>: Attending SIG for information gathering.

<u>Moffitt</u>: Opening a new proteomics facility and is very interested in the Fox Chase LIMS.

Other discussion items:	**Mission Statement** Michael Ochs from Fox Chase will draft a Mission Statement for this SIG. This draft will be circulated to the group for comments. **Regular meeting schedule** The group will meet on the second Tuesday of each month at 2:00 Eastern. Ongoing frequency will be revisited if necessary. **Notes and additional information** Meeting minutes will be emailed and posted on the caBIG on-line forum after review by the group: http://ncicbforums.nci.nih.gov/forums/cabigforum/lfs/icrlfs/SIGs/proteomics Developers should also post additional information about their tools here.

Action Items:	Name Responsible	Action Item	Date Due	Notes
	Michael Ochs	Draft Proteomics SIG mission statement	5/28/04 (approx)	
	Michael Ochs	Post most current proteomics LIMS documentation to the on-line forum	5/28/04 (approx)	
	Juli Klemm	Distribute meeting minutes	5/19/04	
	Juli Klemm	Schedule ongoing meetings	5/19/04	